



(19) BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENTAMT

(12) **Offenlegungsschrift**  
(10) **DE 195 25 282 A 1**

(61) Int. Cl.<sup>8</sup>:  
**C 12 N 15/81**  
C.12 N 9/88

(21) Aktenzeichen: 195 25 282.9  
(22) Anmeldetag: 29. 6. 95  
(43) Offenlegungstag: 2. 1. 97

DE 195 25 282 A 1

(71) Anmelder:

Max-Delbrück-Centrum für Molekulare Medizin,  
13125 Berlin, DE

(72) Erfinder:

Juretzek, Thomas, 03130 Spremberg, DE; Prinz,  
Anke, 34587 Felsberg, DE; Schunck, Wolf-Hagen,  
Dr., 13125 Berlin, DE; Barth, Gerold, Prof. Dr., 01728  
Bannewitz, DE; Mauersberger, Stephan, Dr., 13186  
Berlin, DE

(54) Expressionskassetten zur heterologen Expression von Proteinen in der Hefe *Yarrowia lipolytica* unter Kontrolle des regulierbaren Promotors der Isocitratyase

(57) Die Erfindung hat das Ziel, ein neues, gut regulierbares Expressionssystem für heterologe Gene in der Hefe *Yarrowia lipolytica* zur Verfügung zu stellen. Es soll auf der Basis des regulierbaren Promotors des ICL1 Gens (kodierend für die Isocitratyase) der Hefe *Yarrowia lipolytica* durch eine einfache Kulturführung zur Expression von heterologen Proteinen führen, die unter anderem aufgrund ihrer Eigenschaften zu mikrobiellen Stoffwandlungsprozessen eingesetzt werden können. Der Erfindung liegt die Aufgabe zugrunde, die Hefe *Yarrowia lipolytica* gentechnisch so zu verändern, daß sie bei Kultivierung auf Medien mit geeigneter Kohlenstoffquelle heterologe Proteine exprimiert. Induzierend wirkende Kohlenstoffquellen können sein Ethanol, Acetat, Fettsäuren oder n-Alkane.

Das erfindungsgemäße Verfahren ist dadurch gekennzeichnet, daß Expressionskassetten zur heterologen Expression von Proteinen in *Yarrowia lipolytica*, bestehend aus dem vollständigen Promotor (2186 bp) und Terminator (275 bp) des Isocitratyase Gens (ICL1) und dem zu exprimierenden heterologen Gen (beispielsweise *lacZ*, verschiedene P450 Gene), konstruiert werden. Diese Expressionskassetten werden in einer geeigneten Form in *Yarrowia lipolytica* transformiert und durch Kultivierung der entsprechenden Transformanten dieser Hefe in Medien mit den induzierend wirkenden Kohlenstoffquellen werden die heterologen Proteine durch die Hefe synthetisiert. Anwendungsgebiet der Erfindung ist die Biotechnologie, insbesondere ...

DE 195 25 282 A 1

## Beschreibung

Die Erfindung betrifft ein Verfahren zur Konstruktion von Expressionskassetten, die zur heterologen Expression von Proteinen in der Hefe *Yarrowia* (*Y.*) *lipolytica* unter Kontrolle des regulierbaren Promotors der Isocitratlyase (ICL1 Gen) dieser Hefe geeignet sind.

Sie betrifft ferner die Klonierung des vollständigen ICL1-Promotors, der notwendig ist für die Realisierung des Verfahrens.

Anwendungsgebiet der Erfindung ist die Biotechnologie, vor allem die Nutzung zur Expression von Oligopeptiden und Proteinen in *Y. lipolytica* und insbesondere die sich daraus ergebende Möglichkeit, zielgerichtet mikrobielle Stoffwandlungen mit Hilfe der in der Hefe heterolog exprimierten Enzyme durchzuführen.

Die Isocitratlyase (ICL, EC 4.1.3.1) ist ein Enzym des Glyoxylat-Zyklus, welcher in pro- und eukaryotischen Mikroorganismen, Protozoa, Mollusken, Insekten und Pflanzen nachgewiesen werden konnte. In Säugern ist er dagegen nur in einigen Geweben und nur zu bestimmten Entwicklungsabschnitten nachweisbar (Übersicht bei Vanni et al. 1990 Comp Biochem Physiol 95B: 431-458). Dieser anaplerotische Stoffwechselweg ist notwendig für die Verwertung von Ethanol, Acetat, Fettsäuren und n-Alkanen als Kohlenstoff- und Energiequelle durch Mikroorganismen, wie Bakterien, Hefen und Pilze. Intermediate des Tricarbonsäure-Zyklus (Succinat, Oxalacetat) werden durch den Glyoxylat-Zyklus aus dem im Metabolismus vermehrt entstehenden Acetyl-CoA wieder nachgebildet und in die anabolen Wege (Gluconeogenese, Aminosäuresynthese) eingeschleust.

Heute liegen zur Biochemie und genetischen Regulation des Glyoxylatzzyklus und besonders der ICL in Mikroorganismen (*E. coli*, Hefen wie *Y. lipolytica*, *C. tropicalis*, *S. cerevisiae*, Pilze wie *Aspergillus nidulans*, *Neurospora crassa*) zahlreiche Daten vor. Mutanten in den Strukturgenen der ICL, MS und ACS sowie den regulatorischen Genen wurden isoliert und untersucht. Es gelang unter anderem die Klonierung der ICL-Gene aus den oben angeführten Mikroorganismen (siehe Barth and Scheuber 1993 Mol Gen Genet 241: 422-430).

Die Expression, der Schlüsselenzyme des Glyoxylatzzyklus ICL und Malatsynthese (MS, EC 4.1.3.2) wird in den bisher untersuchten Hefen *S. cerevisiae*, *C. tropicalis* und *Y. lipolytica* auf der transkriptionellen (Induktion/Katabolitrepression) und der posttranslationalen Ebene (Katabolitinaktivierung und metabolische Regulation durch Metaboliten bzw. durch Phosphorylierung/Dephosphorylierung) stark reguliert (Barth 1985 Curr Genet 10: 119-124, Kujau et al. 1992 Yeast 8: 193-203, Barth and Scheuber 1993 Mol Gen Genet 241: 422-430, weitere Literatur siehe Vanni et al. 1990 Comp Biochem Physiol 95B: 431-458).

Es wurde gezeigt, daß Intermediate des Metabolismus (offensichtlich Acetyl-CoA) der induzierend wirkenden C-Quellen notwendig sind für die Induktion der Isocitratlyase. Das induzierend wirkende Acetyl-CoA wird bei Wachstum auf Ethanol, Acetat, Fettsäuren und n-Alkanen verstärkt, dagegen beim Wachstum auf Glucose nur geringfügig gebildet, und die ICL-Synthese wird reprimiert. Glycerol als C-Quelle wirkt dereprimierend auf die ICL. Die stark exprimierte ICL kann bis zu 5% des zellulären Proteins betragen (Vanni et al. 1990 Comp Biochem Physiol 95B: 431-458). Demnach steht im ICL1-Gen offensichtlich ein gut regulierbarer Promotor aus *Y. lipolytica* zur Verfügung, der für die hete-

rologe Expression von Proteinen in diesem Wirt genutzt werden kann.

Seit Klonierung der ersten Gene aus *Y. lipolytica* (LEU2, XPR2) wird diese nicht-konventionelle Hefe auf ihrer Nutzbarkeit zur heterologen Proteinexpression untersucht.

*Y. lipolytica* ist besonders durch ihre ausgeprägte Eigenschaft von praktischem Interesse, eine Reihe von hochmolekularen Proteinen zu sezernieren (Übersicht bei Ogrzydzak 1993 Critical Rev., Biotechnol 13: 1-55), darunter eine alkalische extrazelluläre Protease (AEP), mindestens 3 saure Proteasen, Lipasen bzw. Esterasen, eine Ribonuklease, eine Phosphodiesterase, sowie alkalische und saure Phosphatasen. Das steht im Gegensatz zur Hefe *S. cerevisiae*, die natürlicherweise keine größeren Proteine in Medium ausscheidet.

Dadurch ist die Hefe *Y. lipolytica* (neben weiteren nicht-konventionellen Hefen wie *Pichia pastoris* und *Kluyveromyces lactis*) von potentiell Interesse für ihre Anwendung zur extrazellulären Produktion von heterologen Proteinen (Heslot 1990 Adv Biochem Eng/ Biotechnol 43: 43-73; Buckholz and Gleeson 1991 Bio/Technology 9: 1067-1072; Romanos et al. 1992 Yeast 8: 423-488; Ogrzydzak 1993 Critical Rev Biotechnol 13: 1-55; Sudbery 1994 Yeast 10: 1707-1726).

Bisher wurden in *Y. lipolytica* eine Reihe heterologer Proteine extra- und intrazellulär exprimiert, darunter bakterielle Proteine, wie die  $\beta$ -Galactosidase (*lacZ* Gen) und die  $\beta$ -Glucuronidase (*gusA*) aus *E. coli* unter Kontrolle der LEU2-, LYSS-, XPR2- und PHO2-Promotoren, sowie das Hefeprotein Invertase (*SUC2* Gen) aus *S. cerevisiae* unter Kontrolle des XPR2-Promotors (Literaturangaben in den aufgeführten Übersichtsartikeln).

Von besonderem Interesse war die Expression und Sekretion einer Reihe Proteine höherer Eukaryonten, die auch von kommerziellem Interesse sein können, wie der Blut-Gerinnungsfaktor XIIIa des Menschen (XPR2-Promotor, Tharaud et al. 1992 Gene 121: 111-119), das Prochymosin des Rindes (YPR2 und LEU2-Promotor, Davidow et al. 1987, European Patent Applications EP 220864 B1, and EP 86307839; Franke et al. 1988 Develop Indust Microbiol 29: 43-57; Nicaud et al. 1991 J Biotechnol 19: 259-270), das  $\alpha$ 1-Interferon des Schweines (Heslot et al. 1990 In: Nga BH and Lee YK /Eds/, Microbiology Applications in Food Biotechnology, Elsevier Science, Amsterdam, pp 27-45; Nicaud et al. 1991 J Biotechnol 19: 259-270), das Anaphylatoxin C5a des Menschen (XPR2-Promotor, Davidow et al. 1987 European Patent Application EP 86307839), der gewebespezifische Plasminogen Aktivator (tPA) des Menschen (XPR2-Promotor, Franke et al. 1988 4th ASM Conf Genet Mol Biol Indust Microorganisms, Bloomington, Indiana, Abstracts p37, sowie das Hepatitis B Virus Oberflächenantigen (preS2-HBsAg) unter Kontrolle des XPR2-Promotors (Hämsa und Chattoo 1994 Gene 143: 165-170).

Neben den aufgeführten Berichten zur Expression und Sekretion von Proteinen sollte die alkanverwertende Hefe *Y. lipolytica* ein interessanter Wirt für die Expression von membranständigen (ER) Proteinen, insbesondere von Cytochrom P450 Enzymen sein, da sie eine Reihe von spezifischen Eigenschaften einer alkanverwertenden Hefe (Aufnahme, Transport hydrophober Substrate und günstige Bedingungen für den Elektronentransfer zum P450) aufweist, die für die biotechnolo-

gische Nutzung zur Biotransformation kommerziell interessanter Verbindungen von Vorteil sein können.

Wie aus der oben angeführten Aufzählung hervorgeht, wurden für die heterologe Expression von Proteinen in *Y. lipolytica* bisher hauptsächlich der XPR2- und der LEU2-Promotor genutzt. Der XPR2-Promotor ist am besten charakterisiert und zeigt die bisher für *Y. lipolytica* höchste Promotorstärke, insbesondere nach gezielter Modifikation (Blanchin-Roland et al. 1994 Mol Cell Biol 14: 327–338). Der XPR2-Promotor ist jedoch relativ komplex: durch pH und die N- und C-Quellen reguliert, und erfordert die Kultivierung in einem Medium mit viel Pepton. Die Natur des eigentlichen Induktors ist jedoch nicht bekannt (Blanchin-Roland et al. 1994 Mol Cell Biol 14: 327–338). Die Sekretionssignale (in der Pre-Pro-Sequenz) für die extrazelluläre alkalische Protease wurden, wie oben aufgeführt, vor allem zur Produktion und Sekretion einer Reihe von kommerziell interessanten heterologen Proteinen (Übersichten bei Heslot 1990 Adv Biochem Eng/Biotechnol 43: 43–73; Buckholz und Gleeson 1991 Bio/Technology 9: 1067–1072; Romanos et al. 1992 Yeast 8: 423–488; Ogrydziak 1993 Critical Rev Biotechnol 13: 1–55; Sudbery 1994 Yeast 10: 1707–1726) mit *Y. lipolytica* genutzt. Der LEU2-Promotor ist dagegen nicht als stark einzuschätzen (Gaillardin und Ribet 1987 Curr Genet 11: 369–375).

Es gibt deshalb immer noch einen Bedarf an starken, regulierbaren Promotoren für die Hefe *Y. lipolytica*, wofür offensichtlich die stark exprimierten Gene *PYK1* oder *ICL1* (Strick et al. 1992 Gene 118: 65–72, sowie 1994 Gene 140: 141–143, Barth and Scheuber 1993 Mol Gen Genet 241: 422–430) in Frage kommen.

Das Anliegen der Erfindung ist es, mit der Anwendung des *ICL1*-Promotors einen Beitrag zur Erschließung weiterer regulierbarer Promotoren für *Y. lipolytica* zu leisten, die die Stärke dieses Promotors für eine heterologe Expression mit Hilfe des *lacZ* Genes aus *E. coli* einzuschätzen, und dessen Eignung für die heterologe Expression von Cytochrom P450 in *Y. lipolytica* zu testen.

Die Erfindung hat das Ziel, ein neues, gut regulierbares Expressionssystem für heterologe Gene in der Hefe *Yarrowia* (*Y. lipolytica*) zur Verfügung zu stellen. Es soll auf der Basis des regulierbaren Promotors des *ICL1* Genes (kodierend für die Isocitratlyase) der Hefe *Yarrowia lipolytica* durch eine einfache Kulturführung zur Expression von heterologen Proteinen führen, die unter anderem auf Grund ihrer Eigenschaften zu mikrobiellen Stoffwandlungsprozessen eingesetzt werden können. Der Erfindung liegt die Aufgabe zugrunde, die Hefe *Y. lipolytica* durch die Konstruktion geeigneter Expressionskassetten und deren Transformation in die Hefe gentechnisch so zu verändern, daß sie bei Kultivierung auf Medien mit geeigneter Kohlenstoffquelle heterologe Proteine exprimiert.

Diese Aufgabe wird durch die Klonierung des *ICL1*-Genes, insbesondere seines vollständigen Promotors, aus *Y. lipolytica* und die dadurch ermöglichte Konstruktion von neuen Expressionskassetten mit dem regulierbaren Promotor dieses Genes gelöst.

Das erfindungsgemäße Verfahren ist dadurch gekennzeichnet, daß Expressionskassetten zur heterologen Expression von Proteinen in *Y. lipolytica*, bestehend aus dem vollständigen Promotor (2196 bp) und Terminator (275 bp) des Isocitratlyase Genes (*ICL1*) und dem zu exprimierenden heterologen Gen (beispielsweise das *lacZ* Gen aus *E. coli*, P450 Gene oder cDNA unter-

schiedlicher Herkunft), konstruiert werden.

Das geschieht durch Modifikation des 3'-Endes des Promotors des *ICL1*-Genes aus *Y. lipolytica* mittels PCR (vor oder hinter dem Intron im *ICL1* Gen) zur Einführung eines entsprechenden Restriktionsortes und die dadurch durch einfache Klonierung ermöglichte Fusion mit dem zu exprimierenden heterologen Gen und dem daran angeschlossenen *ICL1*-Terminator, bzw. durch den direkten Anschluß des zu exprimierenden heterologen Genes hinter dem T<sub>350</sub>G<sub>361</sub> des *ICL1*-Genes mittels rekombinanter PCR.

Diese neu konstruierten Expressionskassetten werden dann in das Grundgerüst eines in *Y. lipolytica* autonom replizierenden Vektors, beispielsweise pINA237, pYLI131, oder pYLI131D, umkloniert.

Mit Hilfe dieser neu konstruierten Vektoren werden die Expressionskassetten in einer geeigneten Form in entsprechende Rezipientenstämme der Hefe *Y. lipolytica* transformiert, was sowohl durch integrative Transformation in das Genom als auch durch Transformation mit diesen autonom replizierenden Vektoren möglich ist.

Durch Kultivierung der entsprechenden Transformanten der Hefe *Y. lipolytica* in Mineralsalzmedien mit induzierend wirkenden Kohlenstoffquellen werden die Proteine durch die Hefe synthetisiert und intrazellulär angereichert. Induzierend wirkende Kohlenstoffquellen können sein Ethanol, Acetat, Fettsäuren oder n-Alkane.

Kernpunkt der Erfindung ist die Sequenz des neu klonierten vollständigen Promotors der Isocitratlyase der Hefe *Y. lipolytica*. Davon abgeleitet werden Vektoren zur gentechnischen Veränderung von *Y. lipolytica* konstruiert, die, aufbauend auf dem Grundgerüst des Vektors pINA237, Expressionskassetten, bestehend aus dem vollständigen *ICL1*-Promotor (2196 bp), dem zu exprimierenden heterologen Gen und dem *ICL1*-Terminator (275 bp) kloniert in den pBR322 Teil des Vektors, enthalten.

Eine für die Durchführung der Erfindung wichtige Besonderheit besteht darin, daß mit dem vollständigen *ICL1*-Promotor ein durch einfachen Wechsel der C-Quellen gut regulierbarer Promotor genutzt wird, der die heterologe Expression von Proteinen in *Y. lipolytica* durch eine einfache Kulturführung und unter Einsatz von kostengünstigen C-Quellen wie Ethanol oder Acetat gestattet.

Das erfindungsgemäße Verfahren zeichnet sich dadurch aus, daß die konstruierten Expressionskassetten mit dem vollständigen *ICL1*-Promotor die Expression heterologer Proteine in *Y. lipolytica* gewährleisten, deren Anteil am Gesamtprotein mindestens 3% (wie für die  $\beta$ -Galactosidase aus *E. coli* gezeigt) betragen kann. Der *ICL1*-Promotor gehört demnach zu den durch die C-Quelle einfach regulierbaren Promotoren für *Y. lipolytica*, seine Promotorstärke liegt im Bereich des bisher beschriebenen stärksten Promotors für das XPR2 Gen aus dieser Hefe.

Die Erfindung soll nachfolgend durch Ausführungsbeispiele erläutert werden.

#### 1. Verwendete Hefestämme und Vektoren

Donorstämme zur Genklonierung aus *Y. lipolytica*

B204-12C (MATA met6 spo1),

B204-12C-112 (MATA met6 spo1 GPR1),

Rezipientenstämme zur Transformation von *Y. lipolytica*:

B204-12A-213 (M4TB leu2 ura3) und

B512-3 (MATA ic1c leu2 met6).

Alle Stämme entstammen der Stammsammlung von Dr. G. Barth (Biozentrum der Universität Basel/Technische Universität Dresden).

*E. coli*/Y. lipolytica-Shuttle-Vektoren:

pINA237 — Der *E. coli*/Y. lipolytica-Shuttle-Vektor pINA237 trägt neben dem CEN-ARS18 Gen zur autonomen Replikation in Y. lipolytica das homologe LEU2-Gen sowie einen Anteil von pBR322 zur Amplifikation in *E. coli*. Die Kopiezahl dieses Vektors in Y. lipolytica liegt bei 1–3 pro Zelle (Fournier et al. 1993, Proc Natl Acad Sci USA 90: 4912–4916).

pYLI131 — Der Vektor pYLI131 wurde aus einer einer Genbank (konstruiert aus 2–8 kb großen Sau3A-Fragmenten der chromosomalen DNA des Stammes B204-12C-211 ligiert in den BamHI-Restriktionsort des Vektors pINA237) durch Komplementation in der icl1-Mutante B512-3 (MATA ic1c leu2 met6) direkt kloniert (Abb. 1).

pUCEK2L — lacZ am ATG<sub>319</sub> des ICL1 angeschlossen und in pUC118 kloniert (Abb. 4).

Oligonukleotide zur PCR:

Sa2: 5'-ATG GTG TCG ACA AGG AGA TGG CGC CCA ACA G-3'

3SS: 5'-GCT GTT GCA TGC CTG GGT TAG TAC GGG ACA GAT G-3'

5SS: 5'-GGA TAC TGC ATG CTT TGT ATG CTT GGTC A-3'

Oligonukleotid 1: 5'-GCGGCCGCGTCGACGCGG-3'

Oligonukleotid 2: 5'-GATCCCGCGTCGACGCGGCCCAT-3'

## 2. Das ICL1-Gen aus Y. lipolytica im Ausgangsvektor pYLI131

Das ICL1-Gen mit einem nicht vollständigen Promotor (Promotorvariante A-225 bp) liegt in dem 15,75 kb großen Expressionsvektor pYLI131 vor (Abb. 1). Dieser Vektor wurde durch Direktklonierung in Hefe gewonnen, wobei eine icl1-Mutation des Stammes B5 12-3 mit einer im BamHI-Ort des Expressionsvektors pINA237 angelegten Genbank (durch Ligation partiell Sau3A-verdauter DNA-Fragmente des Stammes B204-12C-112 in den BamHI Ort des Vektors) transformiert wurde. Die Selektion der Transformanten auf den Phänotyp Ace<sup>+</sup> (Acetat verwertend) bzw. Eth<sup>+</sup> (Ethanol verwertend) führte schließlich zur direkten Klonierung des Vektors pYLI131 in Y. lipolytica, der ein funktionell aktives ICL1-Gen trägt.

Der Vektor pYLI131 enthält das ICL1-Gen (Abb. 2) innerhalb eines vollständig sequenzierten Bereiches von 2,5 kb am Anfang eines 7,8 kb Y. lipolytica DNA Fragmentes. Ursprünglich war ein Promotorbereich von 544 bp bis zum ATG<sub>319</sub> als funktionell ausreichend angenommen worden (Abb. 2). Das ICL1-Gen besteht weiter aus einem für das ICL-Protein kodierenden offenen Leserahmen (ORF) von 1625 bp; einem angenommenen Terminatorbereich von 275 bp vom Stopcodon TAA (1982 bp) bis zum BamHI-Ort (2254 bp, entspricht 5696 bp in Abb. 1).

Um den funktionellen Terminatorbereich einzugrenzen, wurde aus dem Vektor pYLI131 (vgl. Abb. 1) zum einen das Sall<sub>644</sub>/BamHI<sub>7979</sub>-Fragment (469 bp) und zum anderen das BamHI<sub>7979</sub>-5696-Fragment (2283 bp) isoliert. Der Vektor pINA237 wurde Sall/BamHI geöffnet und mit den isolierten Fragmenten aus dem pYLI131 der Vektor pYLI131A (10,44 kb) konstruiert, der wie pYLI131 die Promotorvariante A enthält. Die Fähigkeit

dieser Vektoren zur funktionellen Expression der ICL und der  $\beta$ -Galactosidase (Abb. 6) unterscheidet sich jedoch nicht.

Damit konnte experimentell gezeigt werden, daß der 5,3 kb große, nichtsequenzierte Bereich Y. lipolytica-DNA, der sich im Vektor pYLI131 vor dem BamHI-Ort bei 5696 bp (zerstörter BamHI Ort bei 374 bp bis BamHI Ort bei 5696 bp) befindet (siehe Abb. 1), keinen Einfluß auf die Termination der Transkription und damit auf die Expression des ICL1- und des lacZ-Genes hat (siehe Abb. 5 und 6).

In der Nucleotidsequenz des ICL1-Genes im Vektor pYLI131 konnten im Bereich des ICL1-Promotors und -Strukturgens konservierte Intronsequenzen von Y. lipolytica gefunden werden, wie sie aus zwei Y. lipolytica Genen (PYK — Pyruvatkinase, SEC14 — Phosphatidylinositol/Phosphatidylcholin-Transferprotein) bekannt waren (Abb. 2 und 3).

## 3. Klonierung des vollständigen ICL1-Promotors aus Yarrowia lipolytica

Für die Klonierung des ICL1-Promotors wurde die chromosomale DNA aus dem Hefestamm Yarrowia lipolytica B204-12C (MATA met6 spol) nach Kultivierung auf YPD-Vollmedium präpariert und nach Inkubation mit verschiedenen Restriktionsendonukleasen (Sall, XhoI und NcoI und in Kombinationen) im Southernblot analysiert. Die zur Detektion der ICL1-Sequenz verwendete Sonde (987 bp) wurde aus dem Vektor pYLI131 als BamHI<sub>31</sub>-KpnI<sub>956</sub>-Fragment gewonnen, welches den Intronbereich und für den N-terminalen Teil des ICL Protein kodierenden Bereich enthielt (vgl. Abb. 2).

Mit XhoI/NcoI-Fragmenten genomischer DNA der Größe 2,7–3,3 kb wurde durch Klonierung in den durch Sall/NcoI-Verdau vorbereiteten Vektor pUCBM21 (Boehringer Mannheim) eine angereicherte Genbank hergestellt. Das Screening dieser angereicherten Genbank durch Koloniehybridisierung erneut mit der BamHI/KpnI-Sonde lieferte mehrere stark hybridisierende Einzelklone. Von diesen wurden die Plasmide isoliert, durch Restriktionsanalyse ihre Identität festgestellt, und das klonierte Fragment wurde vollständig sequenziert. Das Plasmid pUCIPD enthält in einem 2,9 kb Insert den Promotorbereich bis –2176 bp und einen Teil des ICL-Strukturgens bis zum NcoI Ort bei 702 bp (Abb. 2 und 3). Die Sequenz des vollständigen ICL1 Promotors wurde bestimmt und ist in Abb. 3 dargestellt. Die bisher im pYLI131 vorliegende Sequenz der Promotorvariante A beginnend vom Sau3A-Ort bei –225 bp bis zum NcoI-Ort bei 702 bp im Strukturgen stimmt mit der Sequenz des neu klonierten vollständigen ICL1-Promotors vollständig überein.

## 4. Umklonierung der ICL1-Promotorsequenz in den Vektor pINA237 und Konstruktion des vollständigen ICL1-Genes

Zur Nutzung des vollständigen ICL1-Promotors zur homologen und heterologen Expression von Proteinen wurde dieser in den Vektor pINA237 nach geeigneter Modifizierung umkloniert.

Zur Umklonierung der unter Punkt 2 beschriebenen Sequenz in den für die Expression von Proteinen in Y. lipolytica geeigneten Vektor pINA237 wurde dieser modifiziert, um den sonst nicht vorhandenen Spaltort NotI einzufügen. Dazu wurde nach partieller Spaltung

mit der Restriktase SphI und vollständiger Spaltung mit der Restriktase BamHI eine Oligonukleotidsequenz (Oligonukleotide 1 und 2, siehe 1.) durch Ligation eingefügt, wodurch gleichzeitig der SphI-Spaltort zerstört wurde. In den so gewonnenen Vektor pINA237/Not wurde die ICL1-Promotorsequenz BamHI/NotI und die ICL1-Strukturgensequenz als BamHI-Fragment ligiert. Das dabei entstehende Plasmid pYLI131D enthält das vollständige ICL1-Gen (vgl. Abb. 5 und 6), das eine mit dem chromosomal codierten ICL1-Gen vergleichbare Expressionshöhe und Reprimierbarkeit durch Glucose zeigte.

Eine vergleichbare Klonierung der ICL1-Promotorsequenz BamHI/SalI und der ICL1-Strukturgensequenz als BamHI-Fragment ergab den Vektor pYLI131C (vgl. Abb. 5).

## 5. Heterologe Expression der $\beta$ -Galactosidase (lacZ Gen aus *E. coli*) in *Y. lipolytica* durch den ICL1-Promotor

### 5.1 Konstruktion der Expressionsvektoren pIL22, pIL24, pIL25, pIL31 bis 33

Zur funktionellen heterologen Expression in *Y. lipolytica* wurde das lacZ Gen aus *E. coli* mit dem des ICL1-Promotor (Promotorvarianten A, C und D) fusioniert (Abb. 5 und 6).

Dazu wurden der ICL1-Promotor des Vektor pYLI131 durch PCR am 3'S (Primer Sa2 und 3SS) und am 5'S (Primer Sa2 und 5SS) mit einem SphI-Ort modifiziert, um einen Anschluß des lacZ Genes durch Klonierung zu ermöglichen. Diese modifizierten Promotorfragmente wurden kloniert und in die entsprechend vorbereiteten Plasmide pUCEK2L bzw. pUCEK21L umkloniert (Abb. 4A).

Nach Gewinnung der modifizierten Promotoren, wurden zunächst verschiedene Expressionskassetten im pUC118 vorbereitet, die aus den entsprechenden ICL1-Promotorvarianten, dem Reportergen lacZ und dem ICL-Terminator (275 bp) bestanden.

Die Expressionskassetten wurden als BamHI-Fragmente in die entsprechenden Expressionsvektoren pYLI131, bzw. pYLI131A, 131C, und 131D kloniert. Der Klonierungsweg für diese Expressionsvektoren ist am Beispiel pIL25 in der Abb. 4 dargestellt, da der im Verlauf dieser Klonierung hergestellte Vektor pUCEK4L als Ausgangspunkt für weitere Klonierungen benutzt wurde. Für die Klonierung stand der Vektor pUCEK2L zur Verfügung, in dem das lacZ Gen zwischen dem ICL1-Promotor mit Anschluß am ATG<sub>319</sub> und dem ICL Terminator im pUC118 als Expressionskassette (EK2L) vorlag.

Die Vektoren pIL22, pIL24 und pIL25 sind auf der Basis des Plasmids pYLI131 (Promotorvariante A) kloniert worden.

Die Klonierung der Expressionsvektoren pIL31, pIL32 und pIL33 erfolgte auf Grundlage des Plasmids pUCEK4L (lacZ am 3'S angeschlossen) und der Expressionsvektoren pYLI131A, C und D. Bei diesen Vektoren ist der nichtsequenzierte Teil *Y. lipolytica* DNA nach dem BamHI-Ort bei 5696 bp im pYLI131 (siehe Abb. 1 und Abb. 2) eliminiert. Die in diesen Vektoren enthaltenen Expressionskassetten sind in Abb. 5 dargestellt.

### 5.2 Expression der $\beta$ -Galactosidase unter Kontrolle des ICL Promotors

Mit den Plasmiden pIL24 bis pIL33 konnte die heterologe Expression der  $\beta$ -Galactosidase (lacZ) bei Anschluß an verschiedenen Stellen des ICL-Promotors in Transformanten des *Y. lipolytica* Stammes B204-12A-213 gezeigt werden (Abb. 6).

Die Expression durch die verwendeten Promotorvarianten A, C und D ist durch die C-Quellen Glucose und Ethanol unterschiedlich regulierbar und zeigt deutliche Unterschiede in der Expressionshöhe. Die ursprünglich vorhandene Promotorvariante A (225bp) zeigte dabei eine durch Wachstum auf Ethanol induzierbare Expression des lacZ-Genes, angeschlossen sowohl am 5'-splicing site als auch am 3'-splicing site, wobei der Anschluß am 3'S eine etwa verdoppelte Expressionshöhe zeigte (Abb. 6). Ein Anschluß am ATG<sub>319</sub> im Intron führte jedoch zu keiner lacZ-Expression. Die für die Promotorvariante A gefundene relativ geringe Expressionshöhe für lacZ geht einher mit dem Ausbleiben der Repression durch Glucose. Durch die Klonierung des vollständigen ICL1-Promotors und den daraus abgeleiteten Promotorvarianten D (2176bp) und C (880bp) wurden um den Faktor 7,8 bzw. 4,9 höhere Expressionsraten der  $\beta$ -Galactosidase mit einem Maximalwert nach 10-12 h Wachstum auf Ethanol erreicht. Gleichzeitig ist die Glucoserepression wieder nachweisbar. Demnach sind die längeren Promotorvarianten C und D für eine heterologe Expression von Proteinen in *Y. lipolytica* besser geeignet.

Das Auftreten der Glucoserepression in Abhängigkeit von der Länge des eingesetzten Promotors (in C oder D, nicht jedoch in A) wurde auch bei der plasmidcodierten Expression des homologen ICL1 Strukturgenes in der icl1-Defektmutante B512-3 von *Y. lipolytica* gefunden. Die Expressionshöhe der ICL ist jedoch bei diesen Promotorvarianten etwa vergleichbar und durchläuft ein Optimum bei etwa 8-12 h, was auf die komplexen Regulation der ICL-Aktivität auch auf posttranslationaler Ebene zurückzuführen ist.

Ein Vergleich der Daten zur Expression der  $\beta$ -Galactosidase mit dem stärksten vorhandenen und bisher am häufigsten zur heterologen Expression in *Y. lipolytica* verwendeten XPR2-Promotor (Blanchin-Roland et al. 1994 Mol Cell Biol 14: 327-338) ergab, daß vergleichbare Werte mit dem ICL1-Promotor erreichbar sind.

### Abbildungslegenden

#### Abb. 1: Aufbau des Expressionsvektors pYLI131

Das Plasmid trägt innerhalb von 7,8 kb *Y. lipolytica*-DNA den 1625 bp großen ORF für das ICL1-Strukturgen (ICL1), ein 584 bp Fragment des ICL-Promotors (ICLP), inklusive 358 bp Intronbereich (vgl. Abb. 2), und den ICL-Terminator (ICLT; 275bp). Die restlichen 5,3 kb (zerstörter BamHI Ort bei 374 bp bis BamHI Ort bei 5696 bp) dieses DNA Fragmentes (YLDNA) sind nicht sequenziert. Das Plasmid basiert auf dem in *Y. lipolytica* durch die CENARS18 Region autonom replizierenden und für das LEU2-Gen kodierenden Vektor pINA237. Die *Y. lipolytica* DNA mit dem ICL1-Gen wurde als Sau3A Fragment in den BamHI-Ort (bei 374 bp) des pINA237 ligiert.

#### Abb. 2: Prinzipieller Aufbau des ICL1-Genes aus *Y.*

*lipolytica* mit der Hypothese zum Vorhandensein eines Introns

Dargestellt ist der 2,5 kb große sequenzierte Aus-

schnitt aus dem Vektor pYLI131 am Anfang des 7,8 kb Sau3A Fragmentes, das in den BamHI-Ort des Vektor pINA237 kloniert ist (vgl. Abb. 1). Konservierte Intron Sequenzen: 5'S = 5'-splicing site (GTGAGT<sub>2-7</sub>), 3'S = 3'-splicing site (CAG<sub>357-359</sub>); Die zwischen den unterstrichenen Splicing-Orten liegende nur kursiv geschriebene Sequenz TACTAAC<sub>349-356</sub> wird als branching point Sequenz bezeichnet.

Das nach Splicen neu entstehende A<sub>1</sub>T<sub>350</sub>G<sub>361</sub> als AUG codierend für den angenommenen Translationsstart des ICL-Proteins ist hervorgehoben.

Abb. 3: Sequenz des vollständigen ICL1-Promotors

Der Promotor besitzt eine Größe von 2176 bp vom A<sub>1</sub> aus gerechnet und liegt in einem 2882 bp Fragment im Plasmid pUCIPD kloniert vor. Die Sequenz des Promotors ist ab dem Sau3A-Ort bei -225 bp (Promotorvariante A) stromabwärts bis zum NcoI-Ort bei 702 bp vollkommen identisch mit der Sequenz des ICL1-Genes im Vektor pYLI131 bzw. mit der von Barth und Scheuber (1993) Mol Gen Genet 241: 422-430) publizierten Sequenz des ICL1-Strukturgenes.

Die klein geschriebene Sequenz bis -2201 bp stammt aus dem Klonierungsvektor pUCBM21 und enthält die Orte NotI (-2197 bp), KpnI und ApaI. Die für nachfolgende Klonierungen der Promotorvarianten C und D (siehe Abb. 4 und 5) verwendeten Restriktionsorte Sall (-880 bp) und NotI (-2197 bp) sind wie die Orte Sau3A (-225 bp) und BamHI (-32 bp) fett dargestellt. Das postulierte Intron ist eingerahmt, die Splicingsignale sind fett und kursiv dargestellt. Nach erfolgtem Splicen wurde als neuer Translationsstart das A<sub>1</sub>T<sub>350</sub>G<sub>361</sub> (fett und größer dargestellt) entstehen (vgl. Abb. 2).

Abb. 4: Klonierungsschema zur Herstellung des Expressionsvektors pIL25

A: Klonierung der Expressionskassette im pUCEK4L ausgehend von Vektor pUCEK2L. Insertion des Oligonukleotids siehe Oligonukleotide 1 und 2 unter 1. TF = Transformation. PCR mit den Primern Sa2 und 3SS, vgl. 1.

B: Gewinnung des Expressionsvektors pIL25 durch Umklonierung der EK4L in den Hefektor pYLI131.

Abb. 5: Expressionskassetten mit unterschiedlichen Fragmenten des ICL1-Promotors zur heterologen Expression des lacZ Genes aus E. coli in Y. lipolytica

Die Expressionskassetten EK2L, EK3L, EK4L wurden in das Plasmid pYLI131 (resultierend in den entsprechenden Vektoren pIL22, pIL24, pIL25), bzw. die Expressionskassette EK4L wurde in die Vektoren pYLI131A, 131C und 131D (resultierend in pIL31-pIL33) umkloniert, wie in Abb. 4 prinzipiell dargestellt. PA = Promotorvariante A bis -225 bp, PC = Promotorvariante C bis -880 bp, PD = Promotorvariante D bis -2176 bp (2176 bp), I = Intron, lacZ = lacZ-Strukturgen, T = Terminator, N = nichtsequenzierter Bereich Y. lipolytica DNA, SA = Sau3A' B = BamHI, S = Sall (vgl. Abb. 2).

Abb. 6: Der Aufbau des ICL1 Genes aus Yarrowia lipolytica und der Einfluß unterschiedlicher Promotorvarianten auf die heterologe Expression des Reporterproteins  $\beta$ -Galactosidase (lacZ Gen) aus E. coli in dieser Hefe

Die  $\beta$ -Galactosidase Aktivität wurde im zellfreien Extrakt mit o-Nitrophenyl- $\beta$ -D-galactosid (ONPG) als Substrat bestimmt (10' entspricht 130 nmol/min  $\times$  mg Protein) nach Kultivierung der Transformanten des Stammes B204-12A-213 (M47B leu2 ura3) im Minimalmedium YNB (supplementiert mit Uracil) mit den C-Quellen 1% Ethanol (induzierend) bzw. 1% Ethanol in Anwesenheit von 2% Glucose zur Bestimmung der Katabolitrepresseion durch Glucose. Die Werte wurden nach 8-12 Stunden Kultivierung im Maximum der Expression bestimmt.

#### Patentansprüche

1. Expressionskassetten zur heterologen Expression von Proteinen in der Hefe Yarrowia (Y.) lipolytica, dadurch gekennzeichnet, daß sie aus dem vollständigen Promotor des ICL1-Genes von Y. lipolytica, dem zu exprimierenden heterologen Gen und dem Terminator des ICL1 Genes bestehen.
2. Expressionskassette nach Anspruch 1, dadurch gekennzeichnet, daß sie in das Grundgerüst eines in Y. lipolytica autonom replizierenden Vektors, vorzugsweise pINA237, eingebaut wird.
3. Expressionskassette nach Anspruch 1, dadurch gekennzeichnet, daß sie durch integrative Transformation in das Genom von Y. lipolytica eingebaut wird.
4. Expressionskassette nach Anspruch 1 und 3, dadurch gekennzeichnet, daß sie vorzugsweise in den LEU2 Ort des Genoms von Y. lipolytica eingebaut wird.
5. Expressionskassette nach Anspruch 1, dadurch gekennzeichnet, daß sie die vollständige Promotorsequenz des ICL1 Genes aus Y. lipolytica enthält, wie sie in der Abb. 3 in ihrer Nucleotidsequenz aufgeführt ist.
6. Verfahren zum Aufbau der Expressionskassette nach Anspruch 1, dadurch gekennzeichnet, daß der Promotor des ICL1 Genes aus Y. lipolytica durch Klonierung in einem Vektor für E. coli isoliert wird, dieser Promotor nach Modifikation seines 3'-Endes mittels PCR (vor oder hinter dem Intron im ICL1 Gen) mit dem zu exprimierenden heterologen Gen und dem ICL1-Terminator fusioniert wird, bzw. durch rekombinante PCR das zu exprimierende heterologe Gen direkt hinter dem T<sub>350</sub>G<sub>361</sub> des ICL1-Genes fusioniert wird.
7. Verfahren zum Aufbau der Expressionskassette nach Anspruch 6, dadurch gekennzeichnet, daß der vollständige ICL1-Promotor aus einer angereicherten Genbank im Vektor pUCBM21 kloniert wird, die in Form von 2,7-3,3 kb Fragmenten nach einem NcoI/XhoI-Verdau von genomischer DNA aus Y. lipolytica im entsprechend vorbereiteten Vektor pUCBM21 (NcoI/Sall) erhalten wurde.
8. Verfahren zum Aufbau der Expressionskassette nach Anspruch 6, dadurch gekennzeichnet, daß die Expressionskassette in einen Vektor, welcher für die Transformation von Y. lipolytica geeignet ist, umkloniert wird.
9. Verfahren zum Aufbau der Expressionskassette nach Anspruch 6, dadurch gekennzeichnet, daß die Expressionskassette in einen der Vektoren pINA237, pYLI131, pYLI131A, pYLI131C, oder pYLI131D, welche für die Transformation von Y. lipolytica geeignet sind, umkloniert wird.
10. Verfahren zur Verwendung von mit einer Ex-

pressionskassette gemäß Anspruch 1-5 transformierten Hefezellen, dadurch gekennzeichnet, daß die entsprechenden Transformanden der Hefe *Y. lipolytica* in Mineralsalzmedien mit induzierend wirkenden Kohlenstoffquellen (Ethanol, Acetat, 5 Fettsäuren oder n-Alkane) kultiviert und dadurch die heterolog zu exprimierenden Proteine durch die Hefe synthetisiert werden.

Hierzu 7 Seite(n) Zeichnungen

10

15

20

25

30

35

40

45

50

55

60

65

Abb. 1:

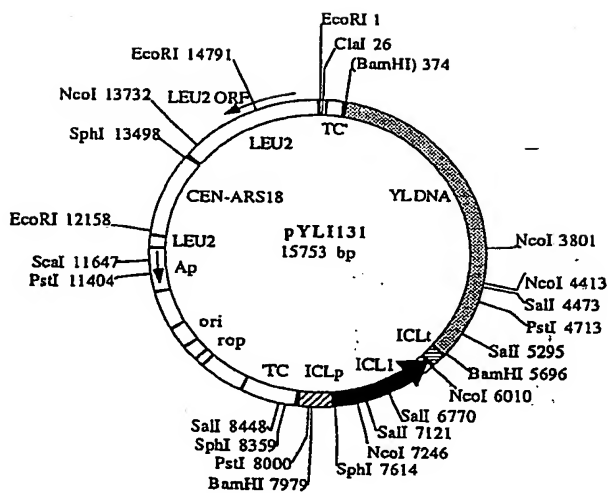




Abb. 2:

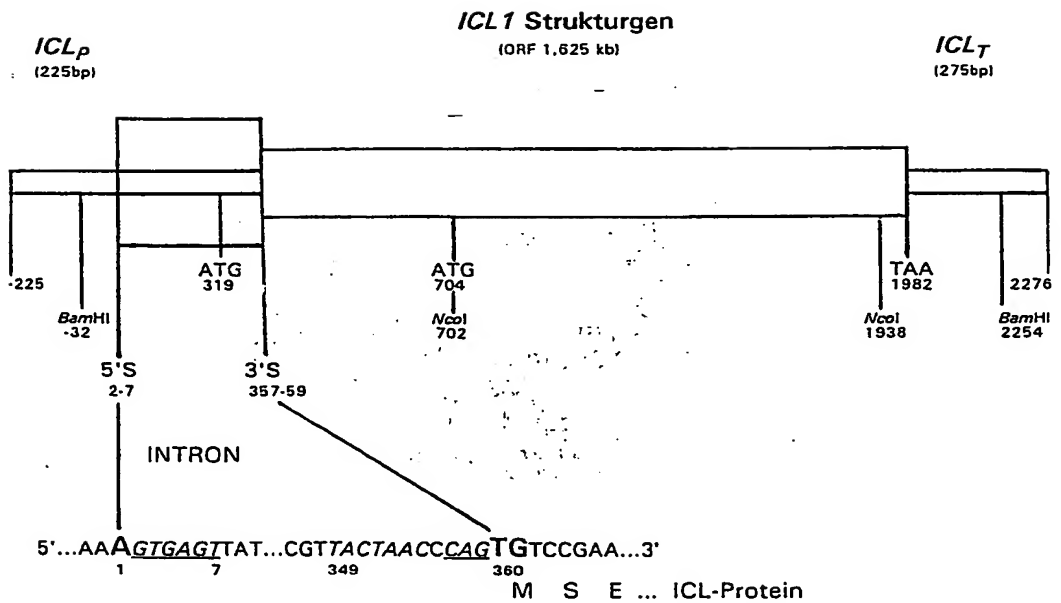


Abb. 3:

-2201 aagtggggcc ggggtaccgg gcccgTCGAG ATGGACATAC TTGTATCGTC GCCCTATGTA ← D  
 -2141 CTCGTAATGC AAGGGATTCC ACCAGACATT CCTGCCACAA TGGCAGGGTC CGTGAAAACG  
 -2081 CCGACCACCTG ACAAGATGCC TTGTTCTGCT TGACCACGGA CTAAGTGCCA CAAGCGAGAT  
 -2021 TAACGTGCTG GGAGACTATT CGGCACACAA GGCCAGACTG TGTTGGCACT CTCATCTCTC  
 -1961 GTACCGACCT CTGTCAACAG TCTAACCGAT TTTTAATGCT CGATATTACC AATGTTTCTT  
 -1901 TGTGCTTTT AACCAAGA ACCGAGCAGA CCCGAACAGG TGCCGAACAT GTGAATAGCA  
 -1841 GTGCTGGAGC TCCATCAGTA AGCATAATAA CACAGCTGCC CAGCGACCTC CAGCCACGGA  
 -1781 CCTCTACCCA GCGACCTCGG GCACGTGACT ATCTGCTCCG TTCTGCGGT CGCTGGCAGG  
 -1721 CTGGCAAATC TGGGGTCTCC ACATTTTCCC CCGGATGTCT TGTTCCGTAG CGTGACTCAT  
 -1661 GCGGAATGAC GTGAATGTAG GAGGGGCTGA GAATGGGGTC CGCAGTTGAT AACCGGGGAT  
 -1601 TATTGGCCGG CGGATTGTCA ACCAGGTGTT TTTACTGGCG TTCTAGAAAT AAAAAGAAAT  
 -1541 AGGCGACCCC CTTGAGCGAG TTCAGCGCGG GCAAAAATGC CTGTTGAAAC ACCTACTTTG  
 -1481 TTCCAGCAC CCCCACGGA TAAATGGAGA CGCATACATC GGCTATGTT GGATACGATC  
 -1421 TTGGGCGGT GTGCGTGGTG TCGCGGTCA TTTTGTCTC CTTTTGGACC CAGCCAGGT  
 -1361 TCAACCGAAC CCCGATTTCG ACATGTGAAA ACGAACAACG GTTTAGTGGG GTTTAAAAAG  
 -1301 TATCAAGTTC AGGGAGGGAA GCATCCAGGC CAACAGCTAT GACCAAGAAA CCAAGCGACC  
 -1241 AAGACATCTG AAGACCAACA AAACCAATAA TCGCTCACC GATGCTCCCC AAACACTAAC  
 -1181 GGCAGACTCT ACTCCAGATT TGCATTGTA GGACCCCGAT ATCGGGTTGC AGATCATGGT  
 -1121 GTCAATCT CTGAACGTGA AGGTTAGGTG GAGGGGATGT TTTGGCCAGA AATGAGCGGT  
 -1061 TTTGTGAGCT TGGAGACGGT AAATCGGATA CGCCAGCGT GAGGATTCCA TAGACCCCTC  
 -1001 CCTTTTGCCA GTATATCCAC CGCAACACCC ACCATGAGCG ACATCTGATA CGGTGCCCGG  
 -941 ACCACTACCC CAAATAAGCT CCAACTAATA TGCCGAGGCA GGTGGGAAAC TATGCACTCC  
 -881 AGTCGACGCT GTAGAAGCAC ATGGAAGGTG CGGAGGCGGT GGCAACGAGG GGCATGAGCC ← C  
 -821 ATCAACGAGT AACCAAGAC AAGCAAGGG GGGAAACGG ACCGGAATCT CTCGCGGTCA  
 -761 CGTGACCCGC CCGGGTTCCA CTCGTCCATG TTGTGCTCT GTGTCTCTCT TCCGACTCGC  
 -701 ATTGGTTAAA CTTCACCCAC CGCAATCACG TCCCACTGGC CAACTTTTT CTGCTTTCTC  
 -641 TGACTTTTTT TGGCCAAAAG GCAACGTCGG AAAGGGTCGG GAGGATTCCG AACCGACGAA  
 -581 AATCGGCCGG CTCAGCGGG GGTAGTTCGG CAGTCTCTGT GGGAGCTCTA GGGGAGCTGT  
 -521 GGTCTGTGTA GGGCCCGGGT CCGGGTTGT TGGGTGTCAA ATCAGGTGT TTTGCCCCC  
 -461 CGCTGAGCG GACTCCGACA ACCGTGTCTC CAACGGCTG ACTAAGTGC TCCGAGCACT  
 -401 CTGCCGTAGC GTTGGTCTGT CTTGTGCGAC TCTGTTCAA GACAGAAGAA AGAAAAAGCT  
 -341 AACCTCCACG TCAGAGACAA TGGTAGAAGG CTGTCTCTT GCAACCGAGG AGAGTGAGTG  
 -281 TTTCTGCGAC GAGGATCATG GGGGTCTGG AGGGTATTT TGAGGGGAAA AAACGGGATC ← A  
 -221 AGGACAAACA GAGGECACAG ACCGGGAATC TGGGCCCCAA AACGGCCTTT TCCCGTCGCA  
 -161 AAACCGGTCT ACATACACCC CTTCGGCCCC CCACAGGCGG GTGTGAAAAA CCCTAAAGCT  
 -101 TGCTTCAAAC CAGACGGACG CACAGCAAGA CACATCATGA AGAGTCACCT GCAGTATATA  
 -41 TAGATCTGGG GATCCCCAGT AGACTGACCA AGCATACAAA AATGAGTAT CCAACAGCGA  
 20 CACGTGAGAT GGCAGAGACA CAGAGACGTG TCTACATGGT TGGACAAGTC TCCACATTCC  
 80 CCAGAGACGT ATCCACATAC AAACACAATC TCACAGCTGA TCTGCTCCTG TGACAGCACA  
 140 GTACATGTTA GTGGATGAGG TGTGTGTAG TGGGTTAAAT GGTGGGACTG ATTCAAGTGGC  
 200 ATCGGTGGCG ACACCTCTA CTCTTCATGT CGTCACCTAC CGTTCGGAAT CCCAATTATC  
 260 TGATGAACTA AACGATTCTT GGCCAAAACA CAATTTTGCC AAAGAAGTCG GTCTACCAAA  
 320 TGCAAGTGTC ACATCAAACA TCTGTCCCGT ACTAACCGAG TGTCCGAACA GCAGCGATTG

I  
N  
T  
R  
O  
N

Abb. 4 A:

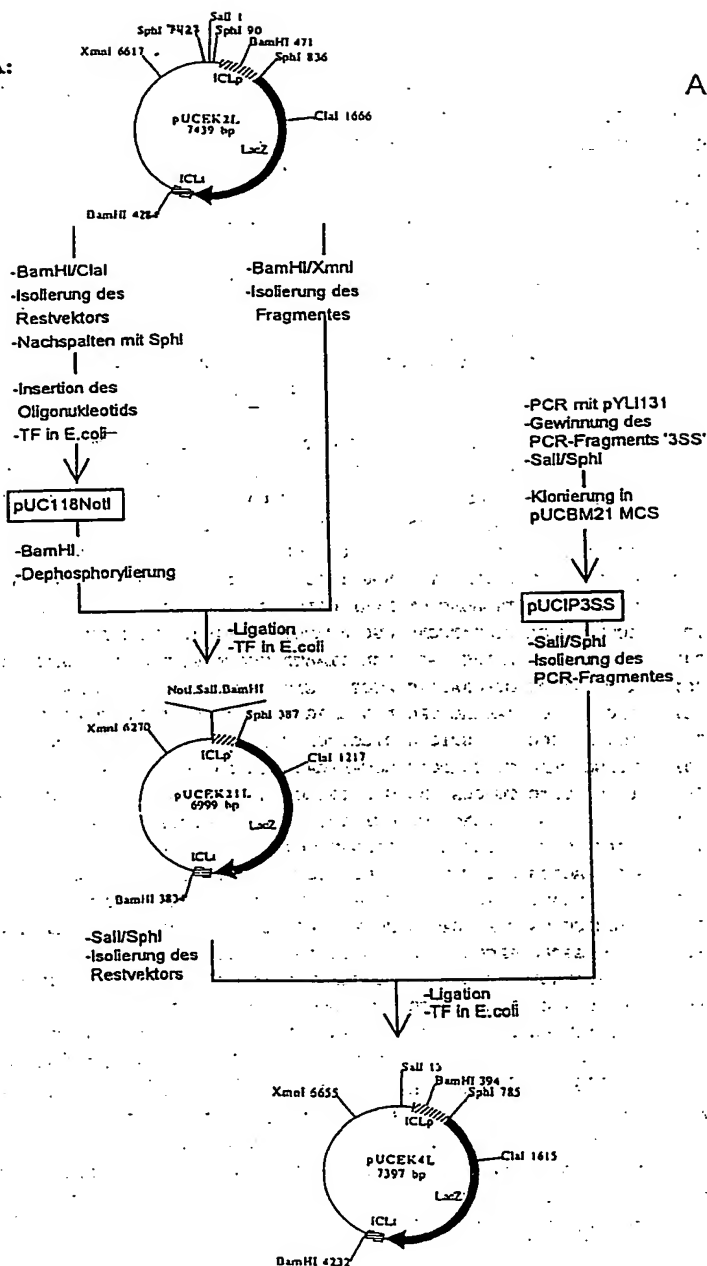


Abb. 4 B:

B

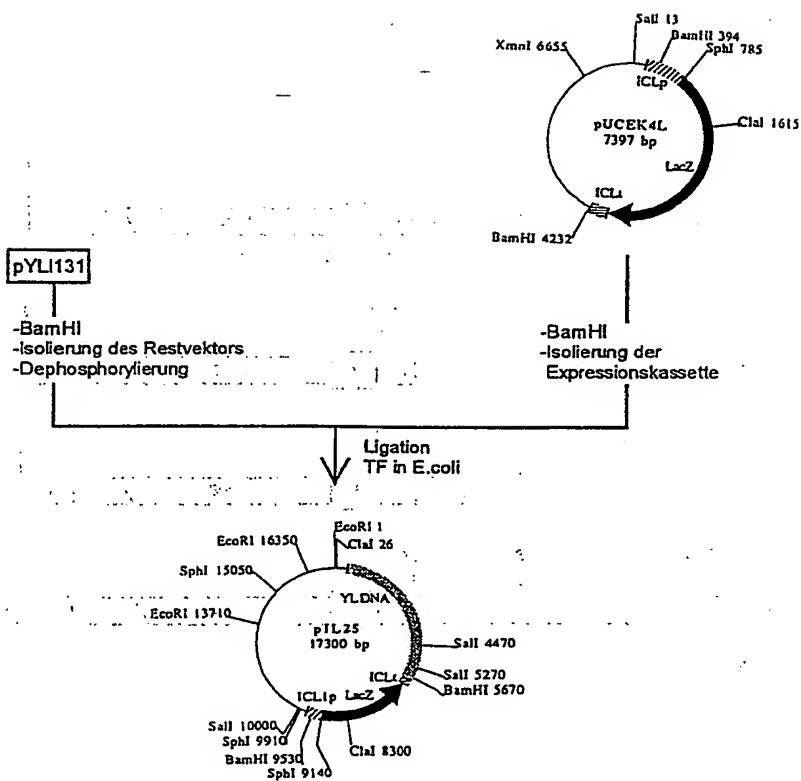


Abb. 5:

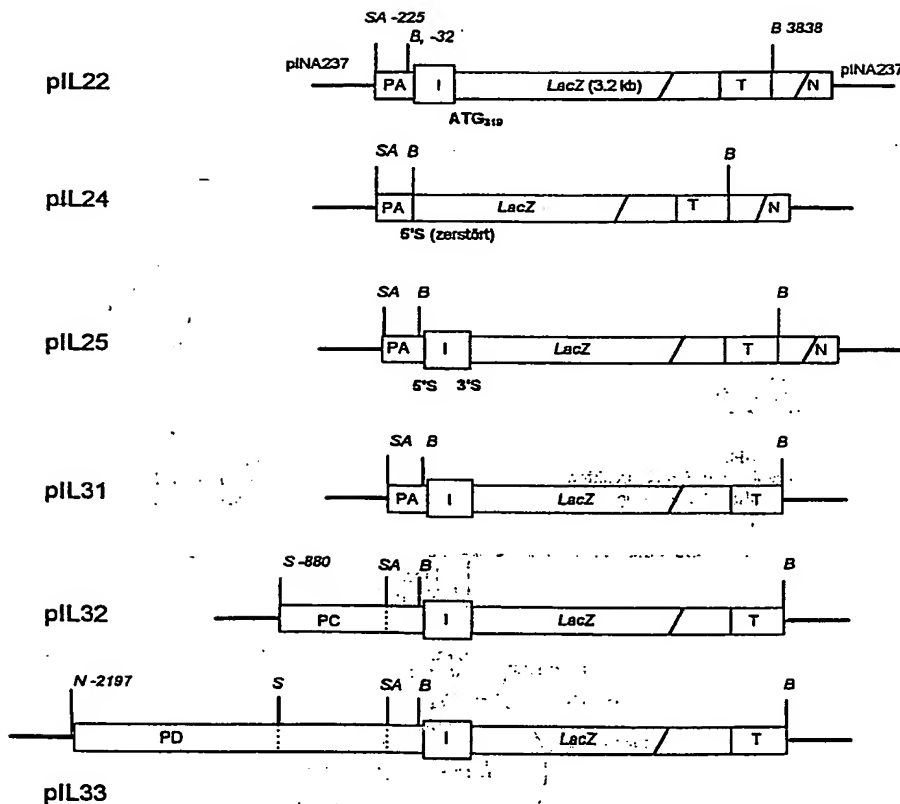


Abb. 6:

